Vugraphs for Appendix A.

OPTIMIZATION WITH RESPECT TO A VECTOR PARAMETER

- Problems in optimization commonly arise involving
  - real and complex vector parameters
  - complex scalar parameters
  - various constraints
- Frequently the quantity to be optimized is *not* analytic
- Need to have effective "power tools" for these problems

GRADIENT WITH RESPECT TO A REAL VECTOR PARAMETER

PROBLEM

Minimize (maximize) the quantity $\mathcal{Q} = \mathcal{Q}(\mathbf{a})$ with respect to the real vector

parameter $\mathbf{a} = \begin{bmatrix} \mathrm{a}_1 \\ \mathrm{a}_2 \\ \vdots \\ \mathrm{a}_N \end{bmatrix}$

APPROACH

Set

$$\nabla_{\mathbf{a}}\mathcal{Q} \stackrel{\text{def}}{=} \begin{bmatrix} \dfrac{\partial \mathcal{Q}}{\partial \mathrm{a}_1} \\ \dfrac{\partial \mathcal{Q}}{\partial \mathrm{a}_2} \\ \vdots \\ \dfrac{\partial \mathcal{Q}}{\partial \mathrm{a}_N} \end{bmatrix} = \mathbf{0}$$

- Will develop formal rules for computing $\nabla_{\mathbf{a}}\mathcal{Q}$.

REAL GRADIENT EXAMPLES

1. For $\quad \mathcal{Q} = \mathbf{b}^T\mathbf{a} = \mathrm{b}_1\mathrm{a}_1 + \mathrm{b}_2\mathrm{a}_2 + \cdots + \mathrm{b}_N\mathrm{a}_N$

$$\nabla_{\mathbf{a}}\mathcal{Q} = \begin{bmatrix} \dfrac{\partial \mathcal{Q}}{\partial \mathrm{a}_1} \\ \dfrac{\partial \mathcal{Q}}{\partial \mathrm{a}_2} \\ \vdots \\ \dfrac{\partial \mathcal{Q}}{\partial \mathrm{a}_N} \end{bmatrix} = \begin{bmatrix} \mathrm{b}_1 \\ \mathrm{b}_2 \\ \vdots \\ \mathrm{b}_N \end{bmatrix} = \mathbf{b}$$

2. For $\quad \mathcal{Q} = \mathbf{a}^T\mathbf{B}\mathbf{a} = \sum_j \sum_k \mathrm{B}_{jk}\,\mathrm{a}_j\,\mathrm{a}_k$

By a similar procedure: $\frac{\partial \mathcal{Q}}{\partial \mathrm{a}_j} = \sum_k (\mathrm{B}_{jk} + \mathrm{B}_{kj})\mathrm{a}_k$

$\Rightarrow \nabla_{\mathbf{a}}\mathcal{Q} = (\mathbf{B} + \mathbf{B}^T)\mathbf{a} \quad (= 2\mathbf{B}\mathbf{a}$ when $\mathbf{B}$ is symmetric)

GRADIENT WITH RESPECT TO A REAL VECTOR PARAMETER FOR SOME COMMON EXPRESSIONS

Quantity $\mathcal{Q}$	$\mathbf{a}^T\mathbf{b}$	$\mathbf{b}^T\mathbf{a}$	$\mathbf{a}^T\mathbf{B}\mathbf{a}$
Gradient $\nabla_{\mathbf{a}}\mathcal{Q}$	$\mathbf{b}$	$\mathbf{b}$	$2\mathbf{B}\mathbf{a}$

Note: $\mathbf{B}$ is assumed to be symmetric.

GRADIENT WITH RESPECT TO A COMPLEX SCALAR QUANTITY

- Functions with dependence $\mathcal{Q} = \mathcal{Q}(\mathrm{a}, \mathrm{a}^*)$ are not analytic; therefore $\dfrac{\partial Q}{\partial \mathrm{a}}$ does not exist.

- If a and a^* are considered separate variables, then partial deriviatives usually exist and are given by

$$\frac{\partial Q}{\partial \mathrm{a}} = \frac{1}{2}\left(\frac{\partial Q}{\partial \mathrm{a_r}} - \jmath \frac{\partial Q}{\partial \mathrm{a_i}}\right) \quad \text{and} \quad \frac{\partial Q}{\partial \mathrm{a}^*} = \frac{1}{2}\left(\frac{\partial Q}{\partial \mathrm{a_r}} + \jmath \frac{\partial Q}{\partial \mathrm{a_i}}\right)$$

GRADIENT WITH RESPECT TO A COMPLEX SCALAR (cont'd.)

- For purposes of optimization, one sets $\dfrac{\partial \mathcal{Q}}{\partial \mathrm{a_r}} = \dfrac{\partial Q}{\partial \mathrm{a_i}} = 0$

- This can be done by defining

$$\nabla_{\mathrm{a}} Q \stackrel{\text{def}}{=} \frac{1}{2}\left(\frac{\partial Q}{\partial \mathrm{a_r}} - \jmath \frac{\partial Q}{\partial \mathrm{a_i}}\right) \quad \text{and} \quad \nabla_{\mathrm{a^*}} Q \stackrel{\text{def}}{=} \frac{1}{2}\left(\frac{\partial Q}{\partial \mathrm{a_r}} + \jmath \frac{\partial Q}{\partial \mathrm{a_i}}\right)$$

and setting *either* $\nabla_{\mathrm{a}} Q$ *or* $\nabla_{\mathrm{a^*}} Q$ to zero.

COMPLEX GRADIENT RELATIONS (SCALAR PARAMETER)

Quantity Q	$\mathrm{a}^*\mathrm{b}$	ab	$\lvert\mathrm{a}\rvert^2 = \mathrm{aa}^*$
Gradient $\nabla_{\mathrm{a}} Q$	0	b	a^*
Gradient $\nabla_{\mathrm{a}^*} Q$	b	0	a

CHECKING FOR A MINIMUM OR MAXIMUM

1. The condition

$$\nabla_{\mathrm{a}} Q = 0 \qquad \textit{or} \qquad \nabla_{\mathrm{a}^*} Q = 0$$

determines a stationary point.

2. For a minimum or maximum, let $\nabla^2_{ab} Q \overset{\text{def}}{=} \nabla_a \Big(\nabla_b Q \Big)$. Then the following *two* conditions must hold:

$$\left(\nabla^2_{aa} Q \right) \cdot \left(\nabla^2_{a^*a^*} Q \right) - \left(\nabla^2_{aa^*} Q \right)^2 < 0$$

and

$$\nabla^2_{aa^*} Q \begin{cases} > 0 & \text{for a minimum} \\ < 0 & \text{for a maximum} \end{cases}$$

COMPLEX GRADIENT ILLUSTRATED (SCALAR PARAMETER)

Find the complex scalar parameter a to minimize

$$\mathcal{Q} = (\mathbf{x} - \mathrm{a}\mathbf{y})^{*T}(\mathbf{x} - \mathrm{a}\mathbf{y})$$

Apply the complex gradient, using results from the table:

$$\begin{aligned}\mathcal{Q} &= \mathbf{x}^{*T}\mathbf{x} - \mathrm{a}^*\mathbf{y}^{*T}\mathbf{x} - \mathrm{a}\mathbf{x}^{*T}\mathbf{y} + |\mathrm{a}|^2\mathbf{y}^{*T}\mathbf{y} \\ \nabla_{\mathrm{a}^*}\mathcal{Q} &= -\mathbf{y}^{*T}\mathbf{x} + \mathrm{a}\,\mathbf{y}^{*T}\mathbf{y} = 0\end{aligned}$$

This yields the result $\boxed{\mathrm{a} = \dfrac{\mathbf{y}^{*T}\mathbf{x}}{\mathbf{y}^{*T}\mathbf{y}}}$

Note: The gradient can also be computed without expanding as

$$\nabla_{\mathrm{a}^*}\mathcal{Q} = -\mathbf{y}^{*T}(\mathbf{x} - \mathrm{a}\mathbf{y})$$

COMPLEX GRADIENT (cont'd.)

The result can be further checked for a minimum. Since

$$\nabla_{\mathrm{a}^*} Q = -\mathbf{y}^{*T}(\mathbf{x} - \mathrm{a}\mathbf{y}) \quad \text{and} \quad \nabla_{\mathrm{a}} Q = -(\mathbf{x} - \mathrm{a}\mathbf{y})^{*T}\mathbf{y}$$

therefore

$$\nabla^2_{\mathrm{aa}^*} Q = \nabla_{\mathrm{a}}\Big(\nabla_{\mathrm{a}^*} Q\Big) = \mathbf{y}^{*T}\mathbf{y} \quad \text{while} \quad \nabla^2_{\mathrm{aa}} Q = \nabla^2_{\mathrm{a}^*\mathrm{a}^*} Q = 0$$

Thus the two conditions for a minimum

$$\left(\nabla^2_{aa} Q\right) \cdot \left(\nabla^2_{a^*a^*} Q\right) - \left(\nabla^2_{aa^*} Q\right)^2 = 0 - (\mathbf{y}^{*T}\mathbf{y})^2 < 0$$

and

$$\nabla^2_{\mathrm{aa}^*} Q = \mathbf{y}^{*T}\mathbf{y} > 0$$

are satisfied.

COMPLEX GRADIENT WITH RESPECT TO A VECTOR PARAMETER

$$\nabla_{\mathbf{a}}\mathcal{Q} = \left(\nabla_{\mathbf{a}^*}\mathcal{Q}^*\right)^* \stackrel{\text{def}}{=} \tfrac{1}{2}\Big(\nabla_{\mathbf{a_r}}\mathcal{Q} - \jmath\nabla_{\mathbf{a_i}}\mathcal{Q}\Big)$$

Quantity $\mathcal{Q}$	$\mathbf{a}^{*T}\mathbf{b}$	$\mathbf{b}^{*T}\mathbf{a}$	$\mathbf{a}^{*T}\mathbf{B}\mathbf{a}$
Gradient $\nabla_{\mathbf{a}}\mathcal{Q}$	$\mathbf{0}$	$\mathbf{b}^*$	$(\mathbf{B}\mathbf{a})^*$
Gradient $\nabla_{\mathbf{a}^*}\mathcal{Q}$	$\mathbf{b}$	$\mathbf{0}$	$\mathbf{B}\mathbf{a}$

Note: $\mathbf{B}$ is assumed Hermitian symmetric.

CONSTRAINED OPTIMIZATION

PROBLEM

Minimize (maximize) the quantity $\mathcal{Q}(\mathbf{a})$ subject to a complex constaint $\mathcal{C}(\mathbf{a}) = 0$.

APPROACH

Form the Lagrangian

$$\boxed{\mathcal{L} = \mathcal{Q}(\mathbf{a}) + \lambda \mathcal{C}(\mathbf{a}) + \lambda^* \mathcal{C}^*(\mathbf{a})}$$

and set and set the *complex gradient* $\nabla_{\mathbf{a}} \mathcal{L}$ or $\nabla_{\mathbf{a}^*} \mathcal{L}$ to zero.

CONSTRAINED OPTIMIZATION (cont'd.)

WHY IT WORKS

Observe that

$$\begin{aligned}
\mathcal{L} &= \mathcal{Q}(\mathbf{a}) + \lambda \mathcal{C}(\mathbf{a}) + \lambda^* \mathcal{C}^*(\mathbf{a}) \\
&= \mathcal{Q}(\mathbf{a}) + 2\mathsf{Re}\, \lambda \mathcal{C}(\mathbf{a}) \\
&= \mathcal{Q}(\mathbf{a}) + 2\lambda_{\mathsf{r}}\, \mathcal{C}_{\mathsf{r}}(\mathbf{a}) - 2\lambda_{\mathsf{i}}\, \mathcal{C}_{\mathsf{i}}(\mathbf{a})
\end{aligned}$$

It is equivalent to adding two *real* constraints, but the first form is more convenient for use with the complex gradient.

Note: When $\mathcal{C}(\mathbf{a})$ is *real*, the last term can be dropped and λ becomes a *real* Lagrange multiplier.

CONSTRAINED OPTIMIZATION ILLUSTRATED

Find $\mathbf{a}$ to maximize

$$\mathcal{Q} = \mathbf{a}^{*T}\mathbf{B}\mathbf{a} \qquad (\mathbf{B} \text{ Hermitian symmetric})$$

subject to the constraint $\mathbf{a}^{*T}\mathbf{a} = 1$.

The constraint is first written as

$$\mathcal{C}(\mathbf{a}) = 1 - \mathbf{a}^{*T}\mathbf{a} = 0$$

where it can be observed that $\mathcal{C}(\mathbf{a})$ is *real*.

CONSTRAINED OPTIMIZATION ILLUSTRATED (cont'd.)

The Lagrangian is

$$\mathcal{L} = \mathbf{a}^{*T}\mathbf{B}\mathbf{a} + \lambda(1 - \mathbf{a}^{*T}\mathbf{a})$$

and the complex gradient condition follows:

$$\nabla_{\mathbf{a}^*}\mathcal{L} = \mathbf{B}\mathbf{a} - \lambda\mathbf{a} = \mathbf{0} \qquad \Longrightarrow \quad \mathbf{B}\mathbf{a} = \lambda\mathbf{a}$$

This shows that $\mathbf{a}$ must be an *eigenvector* of $\mathbf{B}$, but since

$$\mathcal{Q} = \mathbf{a}^{*T}(\mathbf{B}\mathbf{a}) = \mathbf{a}^{*T}(\lambda\mathbf{a}) = \lambda$$

the desired eigenvector to maximize $\mathcal{Q}$ is the one corresponding to the *largest* eigenvalue.